PLOS | ONE

RESEARCH ARTICLE

Method Designed to Respect Molecular Heterogeneity Can Profoundly Correct Present Data Interpretations for Genome-Wide Expression Analysis

Chih-Hao Chen[1,2]*, Chueh-Lin Hsu[1], Shih-Hao Huang[1], Shih-Yuan Chen[1], Yi-Lin Hung[3], Hsiao-Rong Chen[4], Yu-Chung Wu[5,6], Li-Jen Su[1☯]*, H.C. Lee[1,7,8,9☯]*

1 Institute of Systems Biology and Bioinformatics, National Central University, Chungli, Taiwan 32001, **2** Cathay Medical Research Institute, Cathay General Hospital, Taipei, Taiwan 10630, **3** Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan 30013, **4** Department of Medicine, Boston University School of Medicine, Boston, MA 02118, United States of America, **5** Division of Thoracic Surgery, Department of Surgery, Taipei Veterans General Hospital, Taipei, Taiwan 11217, **6** School of Medicine, National Yang-Ming University, Taipei, Taiwan 11221, **7** National Center for Theoretical Sciences, Hsinchu, Taiwan 30043, **8** Department of Physics, National Central University, Chungli, Taiwan 32001, **9** Center for Dynamical Biomarkers and Translational Medicine, National Central University, Chungli, Taiwan 32001

☯ These authors contributed equally to this work.
* chih_hao_chen@yahoo.com (CHC); sulijen@gmail.com (LJS); hclee12345@gmail.com (HCL)

OPEN ACCESS

Citation: Chen C-H, Hsu C-L, Huang S-H, Chen S-Y, Hung Y-L, Chen H-R, et al. (2015) Method Designed to Respect Molecular Heterogeneity Can Profoundly Correct Present Data Interpretations for Genome-Wide Expression Analysis. PLoS ONE 10(3): e0121154. doi:10.1371/journal.pone.0121154

Academic Editor: Zhongxue Chen, Indiana University Bloomington, UNITED STATES

Received: September 5, 2014

Accepted: January 20, 2015

Published: March 20, 2015

Data Availability Statement: All data sets used are public. The URLs for downloading the data sets can be found within the supporting information files.

Funding: This work is partly supported by Grants 98-2627-M-008-004, 99-2911-I-008-100, 100-2911-I-008-001 and 101-2320-B-008-001-MY3 from the Ministry of Science and Technology (ROC), the Cathay General Hospital-NCU Collaboration Grants 99CGH-NCU-A3 and 101-CGH-NCU-B4, and Landseed Hospital- NCU Collaboration Grant NCU-LSH-101-A-008. The funders had no role in study

Abstract

Although genome-wide expression analysis has become a routine tool for gaining insight into molecular mechanisms, extraction of information remains a major challenge. It has been unclear why standard statistical methods, such as the *t*-test and ANOVA, often lead to low levels of reproducibility, how likely applying fold-change cutoffs to enhance reproducibility is to miss key signals, and how adversely using such methods has affected data interpretations. We broadly examined expression data to investigate the reproducibility problem and discovered that molecular heterogeneity, a biological property of genetically different samples, has been improperly handled by the statistical methods. Here we give a mathematical description of the discovery and report the development of a statistical method, named HTA, for better handling molecular heterogeneity. We broadly demonstrate the improved sensitivity and specificity of HTA over the conventional methods and show that using fold-change cutoffs has lost much information. We illustrate the especial usefulness of HTA for heterogeneous diseases, by applying it to existing data sets of schizophrenia, bipolar disorder and Parkinson's disease, and show it can abundantly and reproducibly uncover disease signatures not previously detectable. Based on 156 biological data sets, we estimate that the methodological issue has affected over 96% of expression studies and that HTA can profoundly correct 86% of the affected data interpretations. The methodological advancement can better facilitate systems understandings of biological processes, render biological inferences that are more reliable than they have hitherto been and engender

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

translational medical applications, such as identifying diagnostic biomarkers and drug prediction, which are more robust.

Introduction

Genome-wide expression analysis, based on DNA microarray [1] or the more advanced technology of next-generation sequencing [2], has been a mainstay of genomics research. Its application to discover pathways and functions overrepresented in differentially expressed genes (DEGs) between replicated sample cohorts affords biologists an opportunity to gain holistic insight into molecular mechanisms. For selecting DEGs, the *t*-test was the standard method in the earliest years following the introduction of DNA microarray. Although many studies reported success of application, disparities between results obtained by different groups analyzing similar samples were observed [3–9]. In a later study [10], the MicroArray Quality Control Consortium ascribed the disparities to use of the *t*-test and suggested a hybrid method (HM) employing a non-stringent cutoff for the *p*-value from a *t*-test and a fold-change (FC) cutoff because fold-change ranking was found much more reproducible than *t*-test ranking between platforms and test sites [11, 12]. HM has gained popularity ever since, often applied with a greater-than-1 cutoff for fold-change (on log-2 scale) and a cutoff of 0.05 for the *p*-value from a significance test not limited to the *t*-test. To date, HM and and the *t*-test remain the most adopted methods among few alternatives.

Despite the popularity, we remain concerned about their effectiveness because the reproducibility problem has not been appropriately solved. While the poorer reproducibility of the *t*-test signals compromised specificity, possibly due to its blemished approach for variance estimation, little effort has been made to fully clarify the problem so as to formulate a statistical solution. HM enhances reproducibility but lacks statistical control. It may lose signals for two reasons. One is that continuing use of the *t*-test with a loosened cutoff to lessen its impact on specificity is preposterous and has doubtful effect. The other is that arbitrary cutoffs for fold-change are biased towards selecting genes displaying the most pronounced magnitude of differential expression and may neglect biologically significant signals of unemphatic magnitude. For instance, although a metabolic pathway with all member encoding genes displaying a 20% increase can lead to a vastly higher flux than with a single gene displaying a 20-fold increase, it is far less detectable. The weaknesses can leave pathways or functions falsely prioritized and impact data interpretation.

To investigate the reproducibility problem associated with the *t*-test, we examined 156 expression data sets to research data properties violating principles of the *t*-test. We identified as the problem's primary cause mishandling of molecular heterogeneity, namely the multiplicity of genotypes associated with a phenotype. An accessible example is a patient cohort having two disease subtypes each due to dysfunction of a different pathway. Due to differential expression, the variances of the genes encoding either pathway's members are wider than average. The variances are mistaken for larger error in a *t*-test, leaving the genes deprioritized and the pathways undetected. The cause has wide impact because most expression studies are based on genetically different samples and that none of currently used methods respect molecular heterogeneity. Increasing sample size won't help as it cannot improve gene ranking. The impact has been unnoticed for two reasons. One is the lack of a method to appropriately handling molecular heterogeneity. The other is that conventional assessment of methods has been limited to simulation data, which cannot fully account for biological complexity, and spike-in data (e.g.

the Affymetrix Latin Square data), which are based on genetically identical samples. Using relevant biological data has been desirable but technically unachievable for the inherent DEGs are not priorly known.

Studies of heterogeneous diseases, such as schizophrenia, bipolar disorder and Parkinson's disease, are possibly affected the most by the reproducibility problem. Schizophrenia is a psychiatric disorder that alters basic brain processes of perception, emotion and judgment. Bipolar disorder is a psychiatric disorder that manifests in the form of extreme shifts in a person's mood, energy, and ability to function. Parkinson's disease is a degenerative neurological disorder characterized by impaired dopaminergic transmission. Although their causes remain elusive, the many genome-wide association studies conducted in recent years, which are mostly collected and meta-analyzed at the SZGene [13], BDGene [14] and PDGene [15] databases, have rendered reliable lists of predisposing genes and have provided formidable insight linking the disease etiologies to genetic background. Sporadic cases of Parkinson's disease, which represent the vast majority of all diagnosed cases, have been recognized to have a multi-factorial nature. To date only a restricted number of mechanisms are known to contribute to nigral cell death and mitochondrial dysfunction is among the most well-studied. Since the first link between mitochondria and the disease became evident in the early 1980s, a large body of evidence has accrued to confirm that complex I defect plays a central role [16]. Studies examining gene expression profiles in post-mortem human brain samples from patients compared with healthy controls, on the other hand, have only rendered short lists of overall discordant findings [17–32]. Although the disparities raised concerns, they were often attributed to methodological differences in sample preparation, choice of platform, small sample sizes, and lack of control for factors such as age, brain pH and data quality.

In this article, we explain the reproducibility problem in mathematical terms and present a novel method, named Heterogeneity-corrected Transcriptome Analysis (HTA), for appropriately solving it. We also put forth a novel platform, named Biological Measures Of Relative Reliability (BMORR), for assessing methods using any relevant biological data without priorly knowing the inherent DEGs. On BMORR we comprehensively demonstrate the improved reliability of HTA over conventional methods, using 25 data sets for studying schizophrenia, bipolar disorder and Parkinson's disease, and show the mentioned disparities among previous findings were due to methodological flaws. Based on the 156 data sets, we give an impact assessment of HTA in broadness and profoundness.

Materials and Methods

Data collection and data analysis

The data sets for investigating the reproducibility problem (S1 Table) were randomly collected from the Gene Expression Omnibus (GEO) database and were all produced based on the Affymetrix Human Genome U133 Plus 2.0 platform for relevant biological studies. From the data sets 304 contrasts between replicated cohorts were made in the original studies. The data sets for studying the three diseases (S2 Table) were collected from the Stanley Medical Research Institute, the Harvard Brain Tissue Resource Center and the GEO. All data sets were collected in the Affymetrix CEL formats and analyzed using our own developed software.

For the calculations with HTA, we normalized data using scaling normalization, which scales all arrays' intensities to a global geometric mean; for otherwise calculations, we applied Robust Multi-array Average, which has been the gold standard. We applied no background correction and no control for data quality or confounding factors. We used the Benjamini-Hochberg method [33] to derive false discovery rate (FDR). Our functional analysis was based on the Fisher exact test; the significance level for overrepresentation was $p < 0.001$.

We have also collected from the literature signal lists of the three diseases (S3 Table) for demonstrating our methodological improvements. Note some lists were derived with complicated control for data quality or confounding factors, or through meta-analysis which supposedly improves statistical power.

An explanation of the reproducibility problem

Of the reproducibility problem, we identified limitations by molecular heterogeneity and by sample size as the primary and secondary causes, respectively. For explanation, we categorize samples as type I or type II and divide variance into error and non-error. Type I samples are genetically identical, while type II samples are not. Of the 304 contrasts, 89 are type I. Error is independent of differential expression, is probabilistic and typically follows a normal distribution. Non-error, as explained below, exists only in type II data, arises from differential expression and should not factor into the significance testing. In a *t*-test, sample variance is the estimator of error variance. The estimator is marred by molecular heterogeneity, which manifests itself in type II data as expansion of non-error with absolute fold-change (Fig. 1a and 1b, see more examples in S1 Fig.). Magnitude of the expansion can be pronounced. Although the expansion ostensibly signals variance heterogeneity and justifies variance estimation on a gene-by-gene basis, it results from differential expression and invalidates any method mistaking the affected variances for error and deprioritizing the genes. Increasing sample size won't help. Accuracy of the estimator is further limited by sample size. We illustrate the impact by comparing data of a type I quadruplicate cohort to four simulated arrays, generated by adding Gaussian noise of the same variance to a common template array, in distribution of sample standard deviations (Fig. 1c). The agreement between the two distributions suggests the probe sets share a common error variance, while the data distribution width reveals how easily *t*-test ranking can be disarranged by chance.

A reasonable solution to factoring non-error out of the significance testing is to assume error variance homogeneity among genes and to estimate the common variances based on data of non-DEGs. Such a solution can also mitigate the sample size limitation because the common variances are estimated based on data of many genes.

The HTA method

HTA was devised following the above guideline. It takes error variance as homogeneous among genes and, to better handle samples of uneven quality, heterogeneous among replicates. It assumes most genes are non-DEGs and estimates the samplewise error variances based on data of all genes. Accuracy of the estimation allows the significance be tested using *z*-statistics.

HTA is illustrated in Fig. 2, where it evaluates differential expression of a gene between a test cohort $\{t_i|i = 1,2,3\}$ and a control cohort $\{c_i|i = 1,2,3\}$ in the following steps. (i) HTA estimates the samplewise error variances by pairwisely comparing arrays of a cohort (Fig. 2a). The estimation procedure follows. Let $\{r_i|i = 1,\ldots,n\}$ be the n arrays of a general cohort r. HTA calculates the log-intensity difference of each gene between r_i and r_j and then calculates the variance of the differences, which is denoted by $\sigma^2_{r_i,r_j}$. Assuming errors are normally distributed, we get $\sigma^2_{r_i,r_j} = \sigma^2_{r_i} + \sigma^2_{r_j}$, where $\sigma^2_{r_i}$ and $\sigma^2_{r_j}$ are the samplewise error variances of r_i and r_j, respectively. By taking $\sigma^2_{r_i,r_j}/2$ as an estimate of $\sigma^2_{r_i}$ and taking $\sigma^2_{r_i}$ to be the average of all of its estimates, we get $\sigma^2_{r_i} = 2^{-1}(n-1)^{-1}\sum_{j\neq i}\sigma^2_{r_i,r_j}$. (ii) HTA assigns a Gaussian distribution function (Gaussian) to each measurement of log-intensity (Fig. 2b), taking the measured value as mean and the samplewise error variance as variance, as the probability density function (PDF) of the measurement's true value. (iii) Following scaling normalization (Fig. 2c), HTA multiplies together

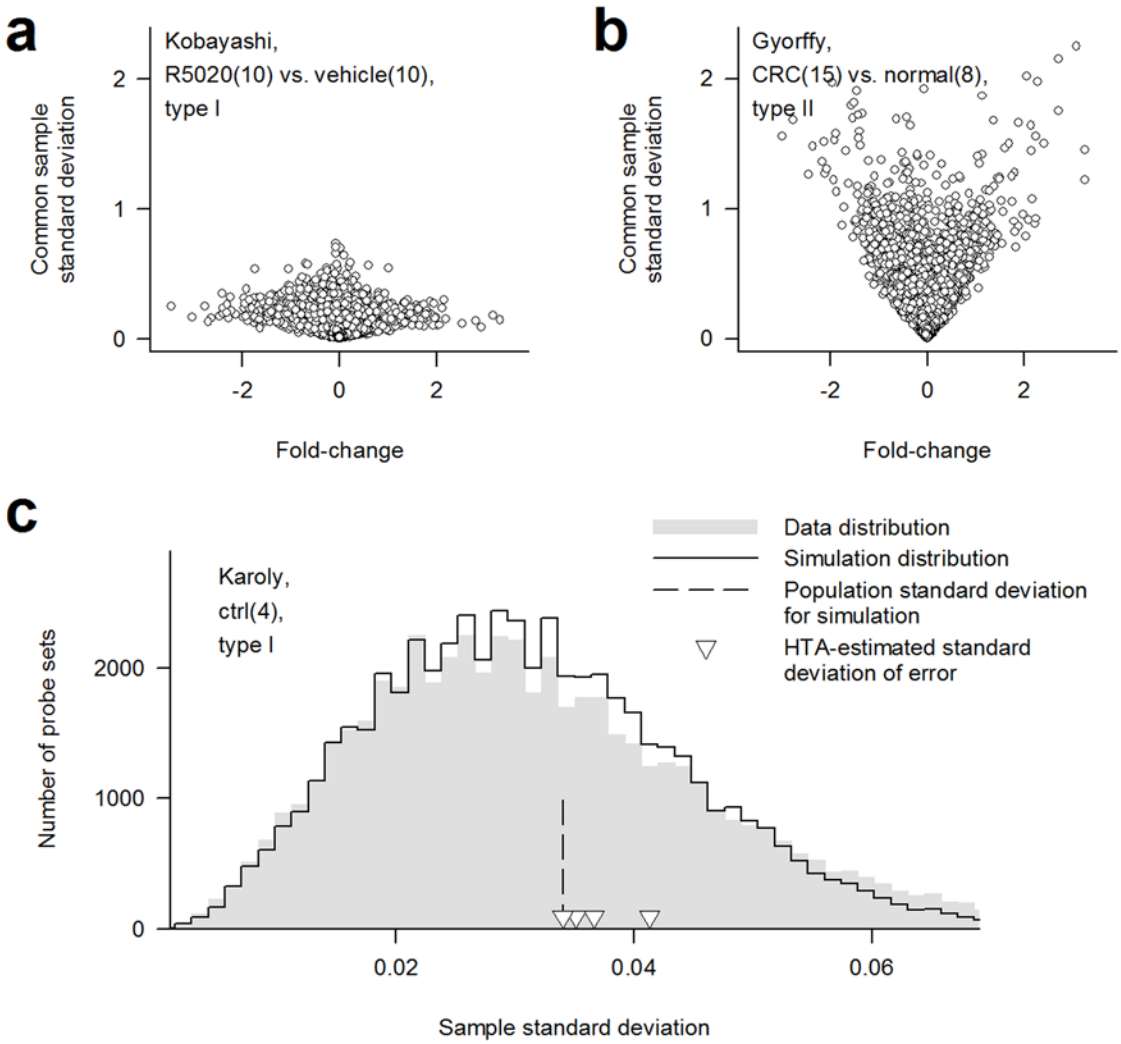

Fig 1. An explanation of the reproducibility problem. (a) Probe set scatter plot of a type I contrast showing independence between sample variance and fold-change. **(b)** Probe set scatter plot of a type II contrast showing dependence between sample variance and fold-change. **(c)** Agreement of distribution of sample standard deviation between a type I contrast and a simulated contrast generated using 0.034 as population standard deviation. Also shown for comparison are the samplewise standard deviations of error estimated using HTA. Printed in the panels are the contrasts or the cohort used. Number in parentheses is number of subjects. Common sample standard deviation, a factor of the denominator in the formula for the *t*-statistic, is defined as $\sqrt{((n_1-1)S_1^2+(n_2-1)S_2^2)/(n_1+n_2-2)}$, where n_i and S_i, $i = 1$ or 2, are respectively number of replicates and sample standard deviation of sample cohort i.

doi:10.1371/journal.pone.0121154.g001

the Gaussians of each cohort (Fig. 2d). The resultant Gaussians, $G_t = G(y; \mu_t, \sigma_t^2)$ and $G_c = G(y; \mu_c, \sigma_c^2)$, are the respective PDFs of the true means of the test and the control cohorts [34]. (iv) The fold-change, the difference between the two true means, can then be predicted using $G_{FC} = G(y; \mu_t - \mu_c, \sigma_t^2 + \sigma_c^2)$ as the PDF. Accordingly, HTA takes $z = (\mu_t - \mu_c)/\sqrt{\sigma_t^2 + \sigma_c^2}$ to be the *z*-static for evaluating differential expression of the gene (Fig. 2e).

Other than solving both limitations, HTA has the following distinctive features. (i) Its error variance estimation is not susceptible to normalization and is much more accurate than that of the *t*-test (Fig. 1c). (ii) It relies solely on the *z*-test for selecting genes and hence provides complete statistical control. (iii) The samplewise error variances facilitate weighting of samples;

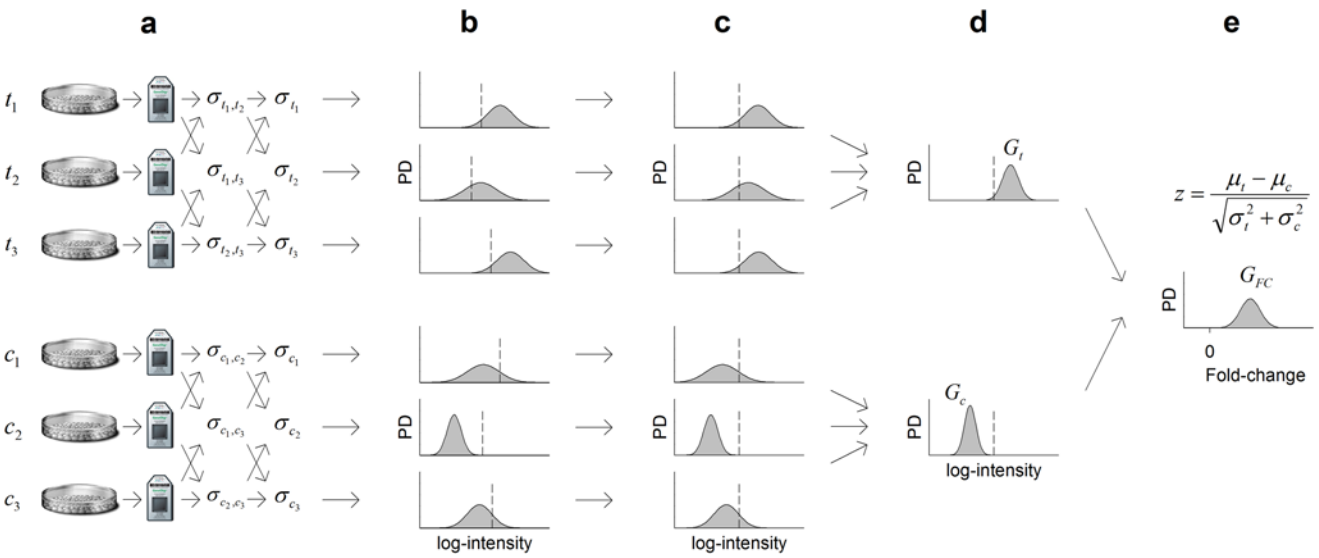

Fig 2. A flow chart of HTA. Here HTA is applied to analyze a gene by contrasting two triplicate cohorts: $\{t_1,t_2,t_3\}$ vs. $\{c_1,c_2,c_3\}$. **(a)** Estimation of samplewise error variances. **(b)** Assigning of a Gaussian to each measurement of log-intensity as the probability density (PD) function of its true value. **(c)** Application of scaling normalization to align the arrays' log-intensity means, marked with the dashed lines. **(d)** Derivation of the PDF of each cohort's mean. **(e)** Derivation of the PDF of the true fold-change and calculation of the *z*-statistic for evaluating differential expression of the gene.

doi:10.1371/journal.pone.0121154.g002

when sample quality is even, HTA ranking is same as fold-change ranking; otherwise, the weighting lessens impact from outliers and makes HTA ranking more favorable.

The BMORR platform

BMORR assesses a method in 3 biological criteria: relative specificity, relative sensitivity and relative reproducibility. The first two are with respect to a single data set. Relative specificity is estimated in number of Gene Ontology functions overrepresented in genes selected under $p < 0.05$, before being divided by that of HTA for normalization. This is because coexpressed genes tend to be functionally coherent but randomly selected genes do not. The *p*-value cutoff ensures the measure is reliably estimated based on sufficient genes even under poor sensitivity. Relative sensitivity is estimated in the product of relative specificity and number of genes selected under a more rigorous cutoff of FDR < 0.05, before being divided by that of HTA for normalization. This is because, under the assumption that relative specificity is proportional to absolute specificity, the product is proportional to number of true positives. Relative reproducibility is with respect to multiple data sets for similar studies and is estimated in average number of times a detected function as described above is repeatedly detected across the data sets, before being divided by that of HTA for normalization.

Results

Validation for BMORR

We used the Kobayashi data set [35] to demonstrate the positive correlation between number of derived functions from probe sets selected under $p < 0.05$ and specificity of the probe sets. The type I data set is a contrast between 10 test samples of human mammary epithelial cells treated with R5020 and 10 controls treated with vehicle. From the data set, we first generated 11 replicates and numbered them from 0 to 10. Next, we permuted the intensities of each array

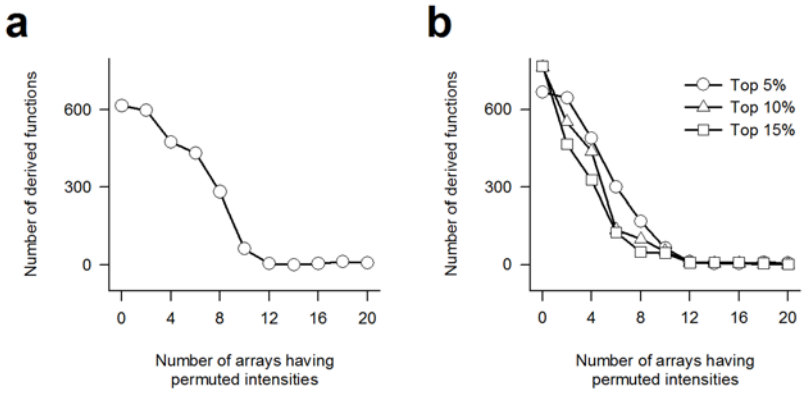

Fig 3. Validation for BMORR. Positive correlation between number of derived functions and specificity demonstrated using **(a)** probe sets selected using the *t*-test($p < 0.05$) and **(b)** top 5%, 10% and 15% probe sets in *t*-test ranking.

doi:10.1371/journal.pone.0121154.g003

of the first *i* test-control pairs of the *i*-th replicate, where $i = 1–10$. We then applied the *t*-test to derive a probe set list from each replicate. The 11 resultant lists supposedly have descending specificities in numerical order. Lastly, we derived functions from the lists and confirmed that they decline with degenerating specificities (Fig. 3a). The decline holds true for top 5%, 10% and 15% probe sets as well (Fig. 3b), indicating the primary cause of the decline is derangement of gene ranking rather than reduction of selected probe sets.

A survey of methods in use

To assess the scale of impact of the reproducibility problem, we surveyed the original studies of the 156 data sets to estimate the adoption rates of the methods in use (Table 1). Only those of 148 data sets have accessible articles which reveal relevant information. The rates were also separately estimated for the 2006–2009 and 2010–2012 periods to identify potential temporal trends. The resultants show no significant change over time. HM (55%) and the *t*-test (26%) were the most popular. All fold-change cutoffs for HM were greater than 1. The methods were followed by GSEA [36] (11%), ANOVA (9%), SAM (8%) and limma [37] (5%). GSEA, SAM

Table 1. Adoption frequencies and rates of data analysis methods.

	2006–2009[1]		2010–2012[2]		Overall	
	Frequency	Rate(%)	Frequency	Rate(%)	Frequency	Rate(%)
t-test	22	25	17	28	39	26
ANOVA	7	8	6	10	13	9
limma	5	6	2	3	7	5
SAM	5	6	7	11	12	8
Other tests[3]	4	5	2	3	6	4
HM	49	56	32	52	81	55
GSEA	10	11	6	10	16	11

[1]Based on 87 data sets. Some studies adopted multiple methods.

[2]Based on 61 data sets. Some studies adopted multiple methods.

[3]Including Wilcoxon rank sum test, Mann-Whitney sum rank test, etc.

doi:10.1371/journal.pone.0121154.t001

and limma represent earlier efforts to solve the reproducibility problem. GSEA bypasses single gene analysis and evaluates data at the level of gene set, namely a group of genes sharing common biological functions, chromosomal location or location; while SAM and limma moderate *t*-statistics by augmenting variances *ad hoc* to keep variances from becoming too small. Collectively, the 6 methods were adopted by 96% of the studies. None of them recognize the problem of overestimated error.

In the following, we present reliability comparisons of HTA to the above methods except ANOVA. ANOVA is not addressed because it is same as the *t*-test when contrasting two sample cohorts.

Comprehensive reliability comparisons on the disease data sets

We compared HTA-derived signals from the 25 contrasts for studying the 3 diseases to those derived using the *t*-test, HM($p < 0.05, |FC| > 1$), limma, SAM and GSEA and to those reported in the literature. The significance test for HM was the *t*-test. The signals were first compared in the following 4 aspects if applicable: (i) number of probe sets selected under FDR < 0.05; (ii) number of functions overrepresented in probe sets selected under $p < 0.05$; (iii) occurrence frequency across the data sets of each disease of the derived functions; (iv) capability of detecting functional signatures of the diseases, measured in bias of derived signals towards selected disease-specific functions. HM was applied with the predetermined cutoffs throughout. For schizophrenia and bipolar disorder, the disease-specific functions were neural functions implicated by both the SZGene and the BDGene lists; for Parkinson's disease, they were mitochondrial functions. For HTA, the *t*-test, HM, limma and SAM, the bias was measured in the *p*-value for overrepresentation; for GSEA, each needed gene set was composed of the platform's probe sets annotated as relevant, the bounds on size of gene set were removed and the bias was measured using the output nominal *p*-value. The derived biases were benchmarked against those expected by chance, derived for validation with the links between probe sets and functions, or between genes and functions, disordered.

Throughout the diseases, HTA rendered far more probe sets under FDR < 0.05 than the *t*-test, HM, SAM and limma (Panels a and d of S2–S5, S8–S11 and S14–S17 Figs.). Regarding overrepresented functions, HTA far outmatched the *t*-test, HM, SAM, limma and the literature-reported signals in number (Panels b and e of S2–S5, S7, S8–S11, S13, S14–S17 and S19 Figs.) and in occurrence frequency (Panels c and f of S2–S5, S7, S8–S11, S13, S14–S17 and S19 Figs.); HTA also far outmatched the *t*-test, HM, SAM, limma, GSEA and the literature-reported signals in coverage of the disease-specific functions (Panels h and k of S2–S5, S7, S8–S11, S13, S14–S17 and S19 Figs.; panels b and e of S6, S12 and S18 Figs.). For schizophrenia and bipolar disorder, HTA detected impairments of the neural functions (S2h and S8h Figs.), which are overrepresented in the SZGene and BDGene lists (S2g and S8g Figs.). For Parkinson's disease, HTA detected mitochondrial dysfunction and pinpointed downregulation of both complex I and ATP synthase complex (S14h Fig.). The problem with complex I is also implicated by the PDGene list (S14g Figs.). The findings were reproducible and as significant as $p = 10^{-20}$. The *t*-test and HM performed poorly in all aspects (S2, S3, S8, S9, S14 and S15 Figs.). SAM and limma rendered slightly more functions than the *t*-test under $p < 0.05$ but zero probe sets under FDR < 0.05 (S4, S5, S10, S11, S16 and S17 Figs.), a reasonable result of global variance augmentation which trades off sensitivity for specificity. GSEA exhibited no sensitivity at all for the functions under discussion (S6, S12 and S18 Figs.). The results of the literature-reported signals (S7, S13 and S19 Figs.) confirmed most of the above findings.

For clearer interpretation, we converted the above results based on BMORR into 4 reliability measures: (i) data set average of relative specificity; (ii) data set average of relative sensitivity;

Table 2. Comprehensive reliability comparisons of HTA to conventional methods and literature-reported signals.

	Schizophrenia				Bipolar disorder				Parkinson's disease			
	RSP[1]	RSE[2]	RR[3]	DRF[4]	RSP[1]	RSE[2]	RR[3]	DRF[4]	RSP[1]	RSE[2]	RR[3]	DRF[4]
HTA	1.00	1.00	1.00	0.60	1.00	1.00	1.00	0.45	1.00	1.00	1.00	0.66
t-test	0.11	0.00	0.05	0.01	0.12	0.00	0.22	0.00	0.19	0.00	0.14	0.21
HM	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.11	0.04
SAM	0.27	0.00	0.25	0.07	0.22	0.00	0.32	0.05	0.33	0.00	0.25	0.29
limma	0.15	0.00	0.07	0.03	0.14	0.00	0.19	0.03	0.29	0.00	0.25	0.27
GSEA	NA	NA	NA	0.00	NA	NA	NA	0.00	NA	NA	NA	0.00
LRS[5]	NA	NA	NA	0.02	NA	NA	NA	0.09	NA	NA	NA	0.07

[1]Average relative specificity compared to HTA.

[2]Average relative sensitivity compared to HTA.

[3]Relative reproducibility compared to HTA.

[4]Average detection rate of the disease-specific functions.

[5]Literature-reported signals.

doi:10.1371/journal.pone.0121154.t002

(iii) relative reproducibility; (iv) data set average of detection rate of the disease-specific functions (Table 2). The results show HTA is remarkably superior in all measures, particularly in sensitivity.

The overall results show the disparities among the previous findings were due to methodological flaws and that a large body of disease information has been overlooked. They also reveal that using the conventional methods rendered the efforts to control for data quality and confounding factors, or to perform meta-analysis, futile.

Because we saw no apparent advantage of SAM, limma and GSEA over the *t*-test, we excluded them from further comparisons.

Broad reliability comparisons and HTA impact assessment

To see if the above results hold true for general studies, we repeated most of the above calculations on the 304 contrasts comparing HTA to the *t*-test and HM (S20–S22 Figs.). Quantitatively, contrast averages of relative specificity and relative sensitivity were respectively 0.36 and 0.20 for the *t*-test, and respectively 0.44 and 0.02 for HM. HTA bettered the *t*-test in relative specificity and relative sensitivity respectively on 98% and 97% of the contrasts, and bettered HM respectively on 89% and 99%. Overall, HTA rendered remarkably improved specificity and sensitivity. The inferior sensitivity of HM revealed $|FC| > 1$ is too stringent, even though it is softer than conventionally chosen.

The broad sensitivity improvement of HTA piqued our curiosity about the amount of lost information in existing data. We estimated that as follows using the 304 contrast. We quantified the content difference between functions derived from a contrast using either the *t*-test (FDR < 0.05) or HM($p < 0.05$,$|FC| > 1$) and those derived using HTA(FDR < 0.05) in area under receiver operating curve (AUC), evaluated taking the latter functions as reference, and considered the contrast has been misanalyzed if AUC < 0.5. Respectively for the *t*-test and HM, we found 74% and 91% of the 304 contrasts have been misanalyzed (Fig. 4). Taking into account the methods' respective adoption rates of 26% and 55%, we conclude HTA can profoundly correct 86% of the affected data interpretations. Note AUC is independent of choice of reference.

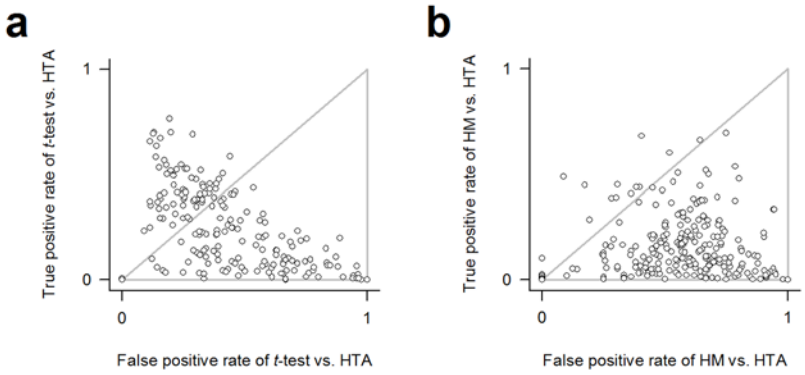

Fig 4. An impact assessment of HTA. In reference to HTA-derived functions from each of the 304 contrasts, we calculated the receiver operating characteristics of **(a)** *t*-test-derived functions and **(b)** HM-derived functions and evaluated the AUC accordingly. The grey triangle in an ROC space encloses the area of AUC < 0.5.

doi:10.1371/journal.pone.0121154.g004

Discussion

We have performed the first study to clarify, solve and broadly appraise the reproducibility problem of genome-wide expression analysis. Our work was based on a total of 328 cohort contrasts derived from 180 data sets produced for relevant biological studies. We present HTA as a solution, elucidate why its simple but rigorous design can solve the two fundamental causes of the problem and demonstrate its improved reliability, comprehensibly and broadly. The demonstration is facilitated by BMORR, a novel platform designed to assess methods using any biological data so that biological complexity, such as molecular heterogeneity, can be taken into account. Using HTA and BMORR we show the problem has affected over 96% of expression studies and that 86% of the affected data interpretations can be profoundly corrected.

HTA is demonstrated on raw data with the simplest normalization strategy, no background correction and no control for data quality or confounding factors. The remarkably improved reliability indicates that mishandling of molecular heterogeneity has been the bottleneck confining the breadth of biomedical research hypotheses explorable and warrants a paradigm shift in future method design. Although the data for the demonstrations were generated using microarrays, molecular heterogeneity as a biological property will equally necessitate HTA, or similarly designed methods, whatever technology is adopted, including next-generation sequencing.

The improved reliability of HTA can benefit a wide spectrum of research fields, ranging from basic biology to the pharmaceutical industry, where it can render inferences that are more reliable than they have hitherto been and engender translational medical applications, such as identifying diagnostic biomarkers and predicting drugs, that are more robust. We also expect HTA to represent an excellent opportunity to rediscover the large body of existing data having been accumulating at public repositories since the introduction of DNA microarray.

Supporting Information

S1 Fig. Manifestation of molecular heterogeneity in expression data. Variance appears independent of fold-change for the type I contrasts on the left; while for the type II contrasts on the right, it tends to expand with absolute fold-change. The contrasts used are as printed. Number

in parentheses is number of subjects. Common sample standard deviation, a factor in the denominator of the formula for the *t*-statistic, is defined as $\sqrt{((n_1-1)S_1^2+(n_2-1)S_2^2)/(n_1+n_2-2)}$, where n_i and S_i, $i = 1$ or 2, are respectively number of replicates and sample standard deviation of sample cohort *i*.
(TIF)

S2 Fig. Superior reliability of HTA over the *t*-test on the schizophrenia data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of *t*-test-derived probe sets. (e) Numbers of *t*-test-derived functions. (f) Occurrence frequency distributions of *t*-test-derived functions. (g) Bias of the SZGene list towards the neural functions. (h) Biases of HTA-derived probe sets towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of *t*-test-derived probe sets towards the neural functions. (k) The biases in (j) expected by chance.
(TIF)

S3 Fig. Superior reliability of HTA over HM on the schizophrenia data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of HM-derived probe sets. (e) Numbers of HM-derived functions. (f) Occurrence frequency distributions of HM-derived functions. (g) Bias of the SZGene list towards the neural functions. (h) Biases of HTA-derived probe sets towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of HM-derived probe sets towards the neural functions. (k) The biases in (j) expected by chance.
(TIF)

S4 Fig. Superior reliability of HTA over SAM on the schizophrenia data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of SAM-derived probe sets. (e) Numbers of SAM-derived functions. (f) Occurrence frequency distributions of SAM-derived functions. (g) Bias of the SZGene list towards the neural functions. (h) Biases of HTA-derived probe sets towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of SAM-derived probe sets towards the neural functions. (k) The biases in (j) expected by chance.
(TIF)

S5 Fig. Superior reliability of HTA over limma on the schizophrenia data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of limma-derived probe sets. (e) Numbers of limma-derived functions. (f) Occurrence frequency distributions of limma-derived functions. (g) Bias of the SZGene list towards the neural functions. (h) Biases of HTA-derived probe sets towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of limma-derived probe sets towards the neural functions. (k) The biases in (j) expected by chance.
(TIF)

S6 Fig. Superior reliability of HTA over GSEA on the schizophrenia data sets. (a) Bias of the SZGene list towards the neural functions. (b) Biases of HTA-derived probe sets towards the neural functions. (c) The biases in (b) expected by chance. (d) GSEA-derived biases towards the neural functions. (e) The biases in (d) expected by chance.
(TIF)

S7 Fig. Superiority of HTA-derived schizophrenia signals over the literature-reported ones. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c)

Occurrence frequency distributions of HTA-derived functions. (d) Numbers of the literature-reported signals. (e) Numbers of functions derived from the literature-reported signals. (f) Occurrence frequency distributions of the functions in (e). (g) Bias of the SZGene list towards the neural functions. (h) Biases of HTA-derived signals towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of the literature-reported signals towards the neural functions. (k) The biases in (j) derived by chance.
(TIF)

S8 Fig. Superior reliability of HTA over the *t*-test on the bipolar disorder data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of *t*-test-derived probe sets. (e) Numbers of *t*-test-derived functions. (f) Occurrence frequency distributions of *t*-test-derived functions. (g) Bias of the BDGene list towards the neural functions. (h) Biases of HTA-derived probe sets towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of *t*-test-derived probe sets towards the neural functions. (k) The biases in (j) expected by chance.
(TIF)

S9 Fig. Superior reliability of HTA over HM on the bipolar disorder data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of HM-derived probe sets. (e) Numbers of HM-derived functions. (f) Occurrence frequency distributions of HM-derived functions. (g) Bias of the BDGene list towards the neural functions. (h) Biases of HTA-derived probe sets towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of HM-derived probe sets towards the neural functions. (k) The biases in (j) expected by chance.
(TIF)

S10 Fig. Superior reliability of HTA over SAM on the bipolar disorder data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of SAM-derived probe sets. (e) Numbers of SAM-derived functions. (f) Occurrence frequency distributions of SAM-derived functions. (g) Bias of the BDGene list towards the neural functions. (h) Biases of HTA-derived probe sets towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of SAM-derived probe sets towards the neural functions. (k) The biases in (j) expected by chance.
(TIF)

S11 Fig. Superior reliability of HTA over limma on the bipolar disorder data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of limma-derived probe sets. (e) Numbers of limma-derived functions. (f) Occurrence frequency distributions of limma-derived functions. (g) Bias of the BDGene list towards the neural functions. (h) Biases of HTA-derived probe sets towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of limma-derived probe sets towards the neural functions. (k) The biases in (j) expected by chance.
(TIF)

S12 Fig. Superior reliability of HTA over GSEA on the bipolar disorder data sets. (a) Bias of the BDGene list towards the neural functions. (b) Biases of HTA-derived probe sets towards the neural functions. (c) The biases in (b) expected by chance. (d) GSEA-derived biases towards the neural functions. (e) The biases in (d) expected by chance.
(TIF)

S13 Fig. Superiority of HTA-derived bipolar disorder signals over the literature-reported ones. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of the literature-reported signals. (e) Numbers of functions derived from the literature-reported signals. (f) Occurrence frequency distributions of the functions in (e). (g) Bias of the BDGene list towards the neural functions. (h) Biases of HTA-derived signals towards the neural functions. (i) The biases in (h) expected by chance. (j) Biases of the literature-reported signals towards the neural functions. (k) The biases in (j) derived by chance.
(TIF)

S14 Fig. Superior reliability of HTA over the *t*-test on the Parkinson's disease data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of *t*-test-derived probe sets. (e) Numbers of *t*-test-derived functions. (f) Occurrence frequency distributions of *t*-test-derived functions. (g) Bias of the PDGene list towards the mitochondrial functions. (h) Biases of HTA-derived probe sets towards the mitochondrial functions. (i) The biases in (h) expected by chance. (j) Biases of *t*-test-derived probe sets towards the mitochondrial functions. (k) The biases in (j) expected by chance.
(TIF)

S15 Fig. Superior reliability of HTA over HM on the Parkinson's disease data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of HM-derived probe sets. (e) Numbers of HM-derived functions. (f) Occurrence frequency distributions of HM-derived functions. (g) Bias of the PDGene list towards the mitochondrial functions. (h) Biases of HTA-derived probe sets towards the mitochondrial functions. (i) The biases in (h) expected by chance. (j) Biases of HM-derived probe sets towards the mitochondrial functions. (k) The biases in (j) expected by chance.
(TIF)

S16 Fig. Superior reliability of HTA over SAM on the Parkinson's disease data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of SAM-derived probe sets. (e) Numbers of SAM-derived functions. (f) Occurrence frequency distributions of SAM-derived functions. (g) Bias of the PDGene list towards the mitochondrial functions. (h) Biases of HTA-derived probe sets towards the mitochondrial functions. (i) The biases in (h) expected by chance. (j) Biases of SAM-derived probe sets towards the mitochondrial functions. (k) The biases in (j) expected by chance.
(TIF)

S17 Fig. Superior reliability of HTA over limma on the Parkinson's disease data sets. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of limma-derived probe sets. (e) Numbers of limma-derived functions. (f) Occurrence frequency distributions of limma-derived functions. (g) Bias of the PDGene list towards the mitochondrial functions. (h) Biases of HTA-derived probe sets towards the mitochondrial functions. (i) The biases in (h) expected by chance. (j) Biases of limma-derived probe sets towards the mitochondrial functions. (k) The biases in (j) expected by chance.
(TIF)

S18 Fig. Superior reliability of HTA over GSEA on the Parkinson's disease data sets. (a) Bias of the PDGene list towards the mitochondrial functions. (b) Biases of HTA-derived probe sets towards the mitochondrial functions. (c) The biases in (b) expected by chance. (d) GSEA-derived biases towards the mitochondrial functions. (e) The biases in (d) expected by chance.
(TIF)

S19 Fig. Superiority of HTA-derived Parkinson's disease signals over the literature-reported ones. (a) Numbers of HTA-derived probe sets. (b) Numbers of HTA-derived functions. (c) Occurrence frequency distributions of HTA-derived functions. (d) Numbers of the literature-reported signals. (e) Numbers of functions derived from the literature-reported signals. (f) Occurrence frequency distributions of the functions in (e). (g) Bias of the PDGene list towards the mitochondrial functions. (h) Biases of HTA-derived signals towards the mitochondrial functions. (i) The biases in (h) expected by chance. (j) Biases of the literature-reported signals towards the mitochondrial functions. (k) The biases in (j) expected by chance.
(TIF)

S20 Fig. Numbers of probe sets and functions derived using HTA($p < 0.05$), *t*-test ($p < 0.05$) and HM($p < 0.05, |FC| > 1$) from the 304 contrasts. Number in parentheses is number of subjects. **(a,c,e,g)** Numbers of probe sets. **(b,d,f,h)** Numbers of functions.
(TIF)

S21 Fig. Numbers of probe sets and functions derived using HTA(FDR < 0.05), *t*-test (FDR < 0.05) and HM($p < 0.05, |FC| > 1$) from the 304 contrasts. Number in parentheses is number of subjects. **(a,c,e,g)** Numbers of probe sets. **(b,d,f,h)** Numbers of functions.
(TIF)

S22 Fig. Superior specificity and sensitivity of HTA over the *t*-test and HM. The 304 contrasts were used. Number in parentheses is number of subjects. (a,c,e,g) Relative sensitivity of the methods. **(b,d,f,h)** Relative specificity of the methods.
(TIF)

S1 Table. The 156 data sets and the 304 contrasts.
(PDF)

S2 Table. The schizophrenia, bipolar disorder and Parkinson's disease data sets.
(PDF)

S3 Table. The literature-reported signals of schizophrenia, bipolar disorder and Parkinson's disease.
(PDF)

S1 Text. The URLs for downloading the used data sets.
(PDF)

Author Contributions

Conceived and designed the experiments: CC. Performed the experiments: CC CH SH SC YH HC. Analyzed the data: CC CH SH SC YH HC. Contributed reagents/materials/analysis tools: CC. Wrote the paper: YW LS HL.

References

1. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary dna microarray. Science 1995; 270: 467–470. doi: 10.1126/science.270.5235.467 PMID: 7569999

2. Shendure J. The beginning of the end for microarrays? Nature Methods 2008; 5: 585–587. doi: 10.1038/nmeth0708-585 PMID: 18587314
3. Tan PK, Downey TJ, Spitznagel EL Jr, Xu P, Fu D, et al. Evaluation of gene expression measurements from commercial microarray platforms. Nucleic acids research 2003; 31: 5676–5684. doi: 10.1093/nar/gkg763 PMID: 14500831
4. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. "stemness": transcriptional profiling of embryonic and adult stem cells. Science 2002; 298: 597–600. doi: 10.1126/science.1072530 PMID: 12228720
5. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, et al. A stem cell molecular signature. Science 2002; 298: 601–604. doi: 10.1126/science.1073823 PMID: 12228721
6. Miller RM, Callahan LM, Casaceli C, Chen L, Kiser GL, et al. Dysregulation of gene expression in the 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine-lesioned mouse substantia nigra. The Journal of neuroscience 2004; 24: 7445–7454. doi: 10.1523/JNEUROSCI.4204-03.2004 PMID: 15329391
7. Fortunel NO, Otu HH, Ng HH, Chen J, Mu X, et al. Comment on "'stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature" (i). Science 2003; 302: 393–393. doi: 10.1126/science.1088249 PMID: 14563990
8. Miklos GLG, Maleszka R. Microarray reality checks in the context of a complex disease. Nature biotechnology 2004; 22: 615–621. doi: 10.1038/nbt965 PMID: 15122300
9. Frantz S. An array of problems. Nature Reviews Drug Discovery 2005; 4: 362–363. doi: 10.1038/nrd1746 PMID: 15902768
10. Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, et al. The microarray quality control (maqc) project shows inter-and intraplatform reproducibility of gene expression measurements. Nature biotechnology 2006; 24: 1151–1161. doi: 10.1038/nbt1239 PMID: 16964229
11. Shi L, Tong W, Fang H, Scherf U, Han J, et al. Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. BMC bioinformatics 2005; 6: S12. doi: 10.1186/1471-2105-6-S2-S12 PMID: 16026597
12. Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, et al. Rat toxicogenomic study reveals analytical consistency across microarray platforms. Nature biotechnology 2006; 24: 1162–1169. doi: 10.1038/nbt1238 PMID: 17061323
13. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the szgene database. Nature genetics 2008; 40: 827–834. doi: 10.1038/ng.171 PMID: 18583979
14. Chang SH, Gao L, Li Z, Zhang WN, Du Y, et al. Bdgene: A genetic database for bipolar disorder and its overlap with schizophrenia and major depressive disorder. Biological psychiatry 2013; 74: 727–733. doi: 10.1016/j.biopsych.2013.04.016 PMID: 23764453
15. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, et al. Comprehensive research synopsis and systematic meta-analyses in parkinson's disease genetics: The pdgene database. PLoS genetics 2012; 8: e1002548. doi: 10.1371/journal.pgen.1002548 PMID: 22438815
16. Blandini F. Neural and immune mechanisms in the pathogenesis of parkinsons disease. Journal of Neuroimmune Pharmacology 2013; 8: 1–13. doi: 10.1007/s11481-013-9435-y
17. Altar CA, Jurata LW, Charles V, Lemire A, Liu P, et al. Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts. Biological psychiatry 2005; 58: 85–96. doi: 10.1016/j.biopsych.2005.03.031 PMID: 16038679
18. Iwamoto K, Bundo M, Kato T. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale dna microarray analysis. Human molecular genetics 2005; 14: 241–253. doi: 10.1093/hmg/ddi022 PMID: 15563509
19. Middleton FA, Mirnics K, Pierri JN, Lewis DA, Levitt P. Gene expression profiling reveals alterations of specific metabolic pathways in schizophrenia. The Journal of neuroscience 2002; 22: 2718–2729. PMID: 11923437
20. Mirnics K, Middleton FA, Marquez A, Lewis DA, Levitt P. Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. Neuron 2000; 28: 53–67. doi: 10.1016/S0896-6273(00)00085-4 PMID: 11086983
21. Arion D, Unger T, Lewis DA, Levitt P, Mirnics K. Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. Biological psychiatry 2007; 62: 711–721. doi: 10.1016/j.biopsych.2006.12.021 PMID: 17568569
22. Hakak Y, Walker JR, Li C, Wong WH, Davis KL, et al. Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proceedings of the National Academy of Sciences 2001; 98: 4746–4751. doi: 10.1073/pnas.081071198

23. Aston C, Jiang L, Sokolov BP. Microarray analysis of postmortem temporal cortex from patients with schizophrenia. Journal of neuroscience research 2004; 77: 858–866. doi: 10.1002/jnr.20208 PMID: 15334603

24. Dracheva S, Davis KL, Chin B, Woo DA, Schmeidler J, et al. Myelin-associated mrna and protein expression deficits in the anterior cingulate cortex and hippocampus in elderly schizophrenia patients. Neurobiology of disease 2006; 21: 531–540. doi: 10.1016/j.nbd.2005.08.012 PMID: 16213148

25. Nakatani N, Hattori E, Ohnishi T, Dean B, Iwayama Y, et al. Genome-wide expression analysis detects eight genes with robust alterations specific to bipolar i disorder: relevance to neuronal network perturbation. Human molecular genetics 2006; 15: 1949–1962. doi: 10.1093/hmg/ddl118 PMID: 16687443

26. Ryan M, Lockstone H, Huffaker S, Wayland M, Webster M, et al. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. Molecular psychiatry 2006; 11: 965–978. doi: 10.1038/sj.mp.4001875 PMID: 16894394

27. Seifuddin F, Pirooznia M, Judy JT, Goes FS, Potash JB, et al. Systematic review of genome-wide gene expression studies of bipolar disorder. BMC psychiatry 2013; 13: 213. doi: 10.1186/1471-244X-13-213 PMID: 23945090

28. Grünblatt E, Mandel S, Jacob-Hirsch J, Zeligson S, Amariglo N, et al. Gene expression profiling of parkinsonian substantia nigra pars compacta; alterations in ubiquitin-proteasome, heat shock protein, iron and oxidative stress regulated proteins, cell adhesion/cellular matrix and vesicle trafficking genes. Journal of neural transmission 2004; 111: 1543–1573. doi: 10.1007/s00702-004-0212-1 PMID: 15455214

29. Hauser MA, Li YJ, Xu H, Noureddine MA, Shao YS, et al. Expression profiling of substantia nigra in parkinson disease, progressive supranuclear palsy, and frontotemporal dementia with parkinsonism. Archives of neurology 2005; 62: 917. doi: 10.1001/archneur.62.6.917 PMID: 15956162

30. Moran L, Duke D, Deprez M, Dexter D, Pearce R, et al. Whole genome expression profiling of the medial and lateral substantia nigra in parkinsons disease. Neurogenetics 2006; 7: 1–11. doi: 10.1007/s10048-005-0020-2 PMID: 16344956

31. Miller RM, Kiser GL, Kaysser-Kranich T, Lockner RJ, Palaniappan C, et al. Robust dysregulation of gene expression in substantia nigra and striatum in parkinson's disease. Neurobiology of disease 2006; 21: 305–313. doi: 10.1016/j.nbd.2005.07.010 PMID: 16143538

32. Lewandowski NM, Ju S, Verbitsky M, Ross B, Geddie ML, et al. Polyamine pathway contributes to the pathogenesis of parkinson disease. Proceedings of the National Academy of Sciences 2010; 107: 16970–16975. doi: 10.1073/pnas.1011751107

33. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological) 1995; 289–300.

34. Chen CH, Lee HC, Ling Q, Chen HR, Ko YA, et al. An all-statistics, high-speed algorithm for the analysis of copy number variation in genomes. Nucleic acids research 2011; 39: e89–e89. doi: 10.1093/nar/gkr137 PMID: 21576227

35. Kobayashi S, Stice JP, Kazmin D, Wittmann BM, Kimbrel EA, et al. Mechanisms of progesterone receptor inhibition of inflammatory responses in cellular models of breast cancer. Molecular Endocrinology 2010; 24: 2292–2302. doi: 10.1210/me.2010-0289 PMID: 20980435

36. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 2005; 102: 15545–15550. doi: 10.1073/pnas.0506580102 PMID: 16199517

37. Smyth GK. Limma: linear models for microarray data. In: Bioinformatics and computational biology solutions using R and Bioconductor, Springer; 2005.